# Pattern-Based Linked Data Publication: The Linked Chess Dataset Case

Víctor Rodríguez-Doncel[1], Adila A. Krisnadhi[2,3], Pascal Hitzler[2], Michelle Cheatham[2], Nazifa Karima[2], and Reihaneh Amini[2]

[1] Ontology Engineering Group, Universidad Politécnica de Madrid
[2] Data Semantics Lab, Wright State University
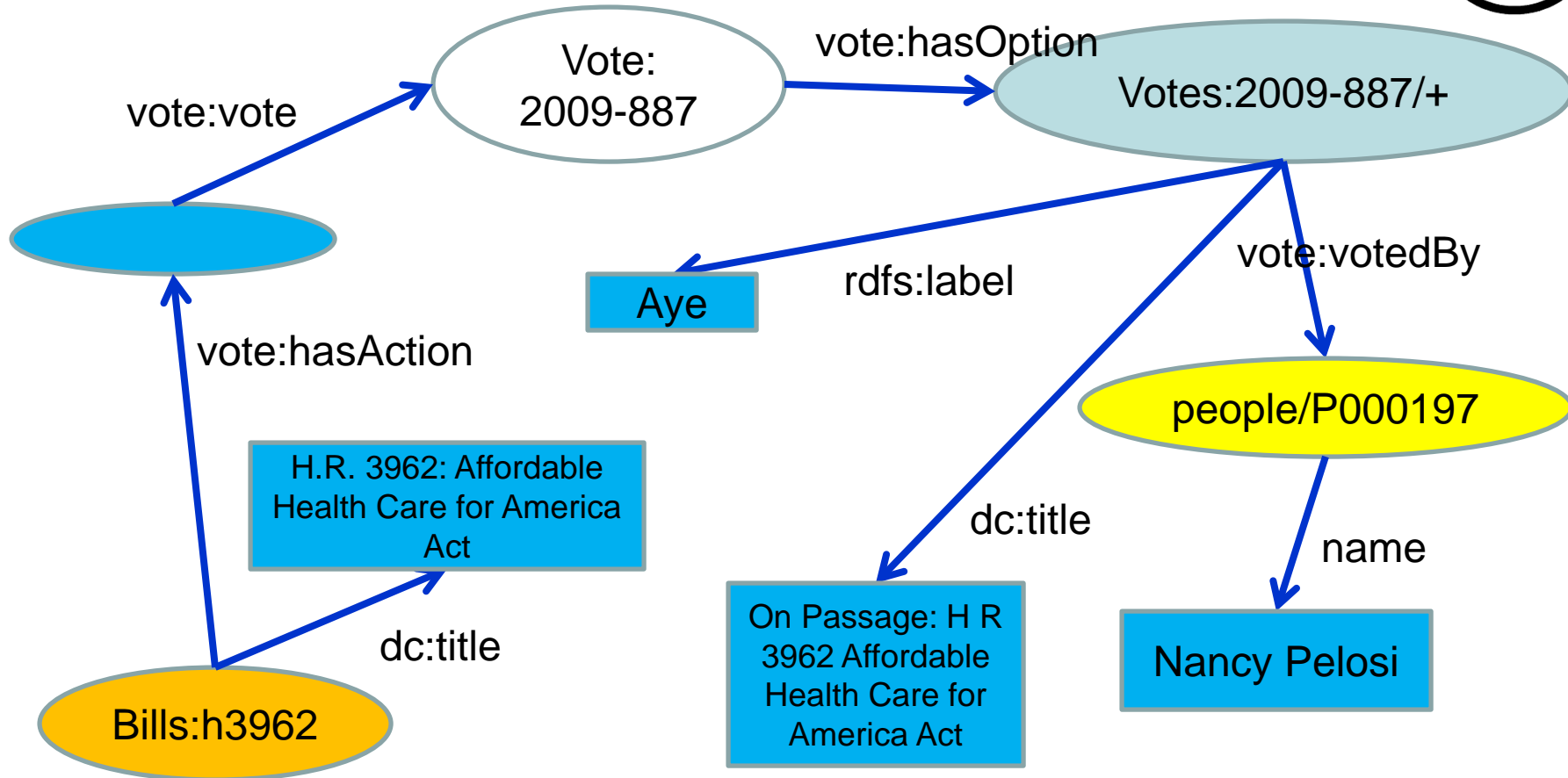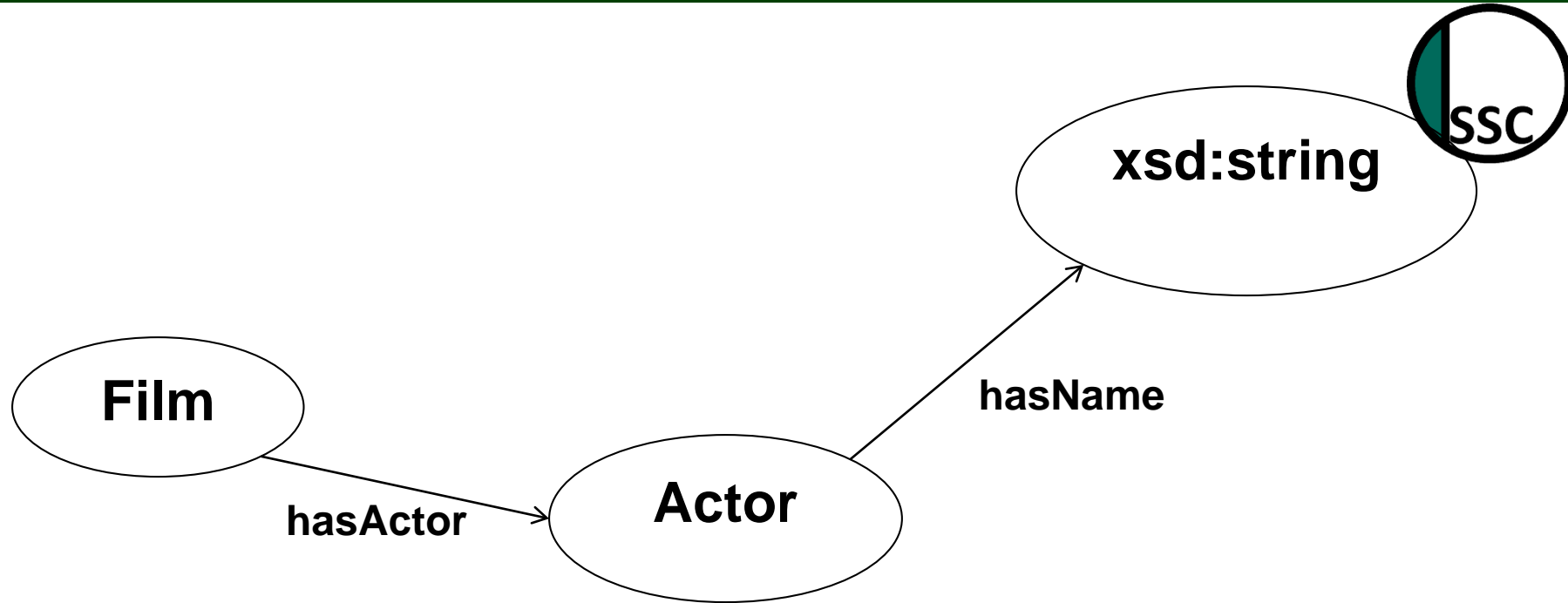[3] Faculty of Computer Science, Universitas Indonesia

# A reviewer's words

"… I suppose if we are still seeing a lack of contributions and applications with respect to consuming Linked Data, one would have to question how Linked Data is published in the first-place. And indeed getting publishers to agree on patterns – not just vocabulary – would seem to make consumers' lives that little bit easier. ... Maybe agreement on patterns – not just vocabulary – is what we need to help kick-start consumption of Linked Data"
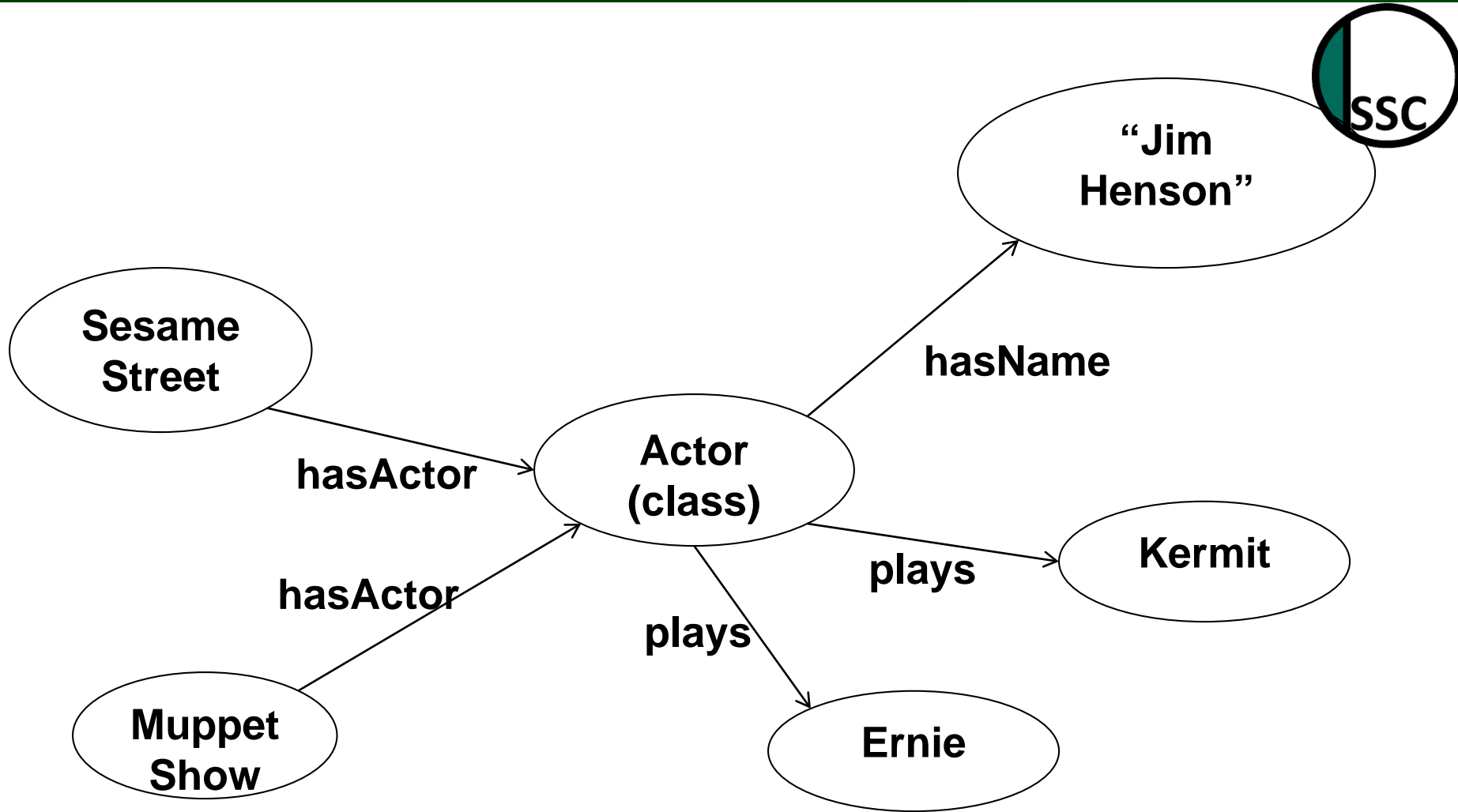
# Reusing Linked Data is tricky

**"Nancy Pelosi voted in favor of the Health Care Bill."**

vote:vote

Vote: 2009-887

vote:hasOption

Votes:2009-887/+

rdfs:label

Aye

vote:votedBy

people/P000197

vote:hasAction

H.R. 3962: Affordable Health Care for America Act

dc:title

dc:title

On Passage: H R 3962 Affordable Health Care for America Act

name

Nancy Pelosi

Bills:h3962

WRIGHT STATE UNIVERSITY

**DaSe Lab**

SSC

"Jim Henson"

Sesame Street

Actor (class)

hasActor

hasActor

Muppet Show

hasName

plays

Kermit

plays

Ernie

# Implicit Ontologies

When publishing Linked Data, there is always an underlying graph schema, which somebody has "designed."

In other words, there is always an underlying ontology
> even if the provider hasn't bothered to write it up properly or share it.

The W3C Shapes Working Group is probably out to make this explicit.

(I'm not sure why they don't call them "ontologies" though, but I can also live with "RDF Shapes" of course.)

# No LOD without ontologies

You can't avoid the schema when dealing with Linked Data.

Which means you also can't avoid the ontology/schema modeling issues.

If your schema is not well-designed and well-documented, then it will not be easily reusable.

[Looking forward to seeing more about the RDF Shapes work.]

# From ontologies to linked data

We recently realized we have a lot of chess players in the lab.
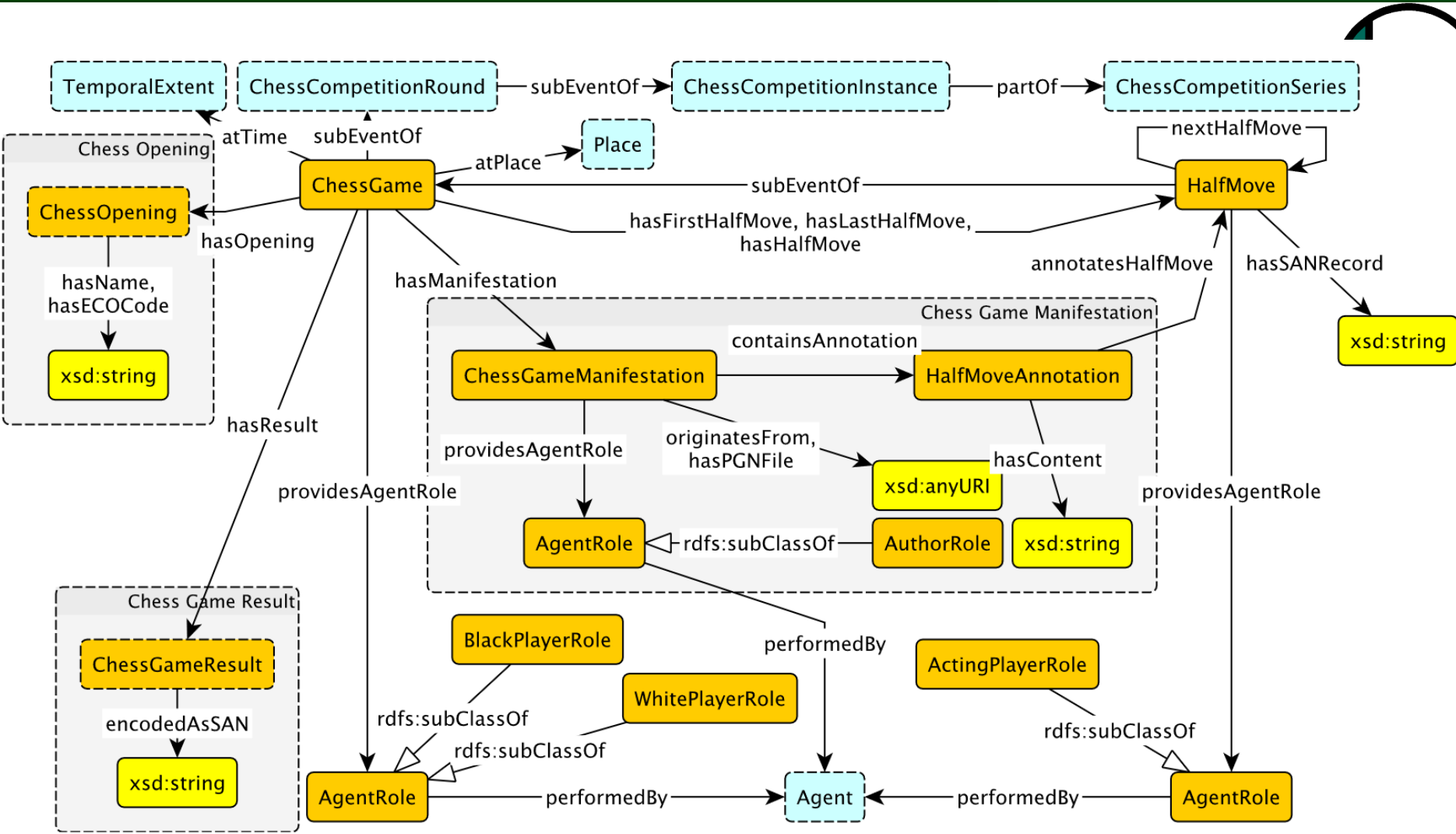
And that there's no linked dataset for chess games.

So we decided to change that.

There is already an established standard, the Portable Games Notation PGN (text-based, with some basic metadata), and lots of data available on the Web.

Following our own recommendations, we first made an ontology …

- **Collaborative modeling**, group ideally has
  - **More than one domain experts.**
  - **People familiar with the base data.**
  - **People understanding possible target use cases.**
  - **An ontology engineer familiar with the modeling approach.**
  - **Somebody who understands formal semantics of OWL.**

- **Domain experts are queried as to the main notions for the application domain.**
  - **E.g. for chess, these would include**
    - **Chess game; move; opening; tournament; players; commentary**

- **From available data and from application use cases, devise <span style="color:red">competency questions</span>, i.e. questions which should be convertible into queries, which in turn should be answerable using the data.**

  <span style="color:green">Retrieve all games where Fischer lost in the poisoned pawn variation of the Sicilian.</span>
  <span style="color:green">Retrieve all games where Fischer opened 1. Nf3.</span>

- **Then <span style="color:red">prioritize</span> which notions to model first. In the chess case, e.g.**

  <span style="color:green">chess game
  move/half-move
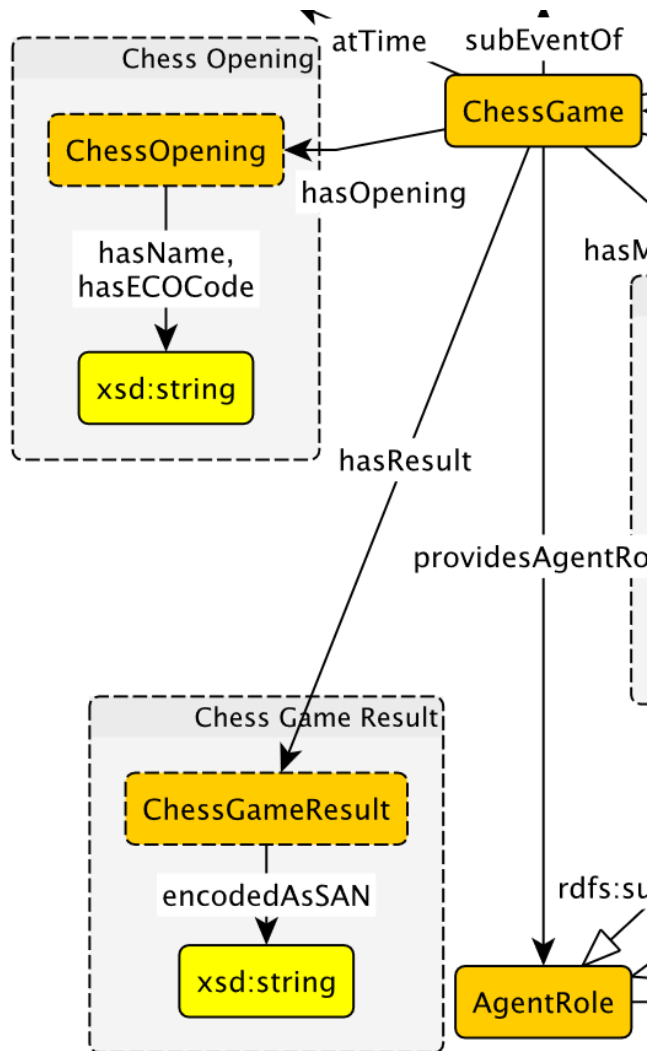  players
  opening
  commentary
  tournaments</span>

# GeoVoCamps modeling approach

- **Understand the nature of the things you are modeling.**

| | | |
|---|---|---|
| **Chess game** | **...** | **An Event** |
| **Half-move** | **…** | **A Subevent of a chess game** |
| **Player** | **…** | **The Role of an Agent** |
| **Opening** | **…** | **this is probably complex** |
| **commentary** | **…** | **this is again more complex** |
| **tournaments** | **…** | **Events** |

# Opening and games result
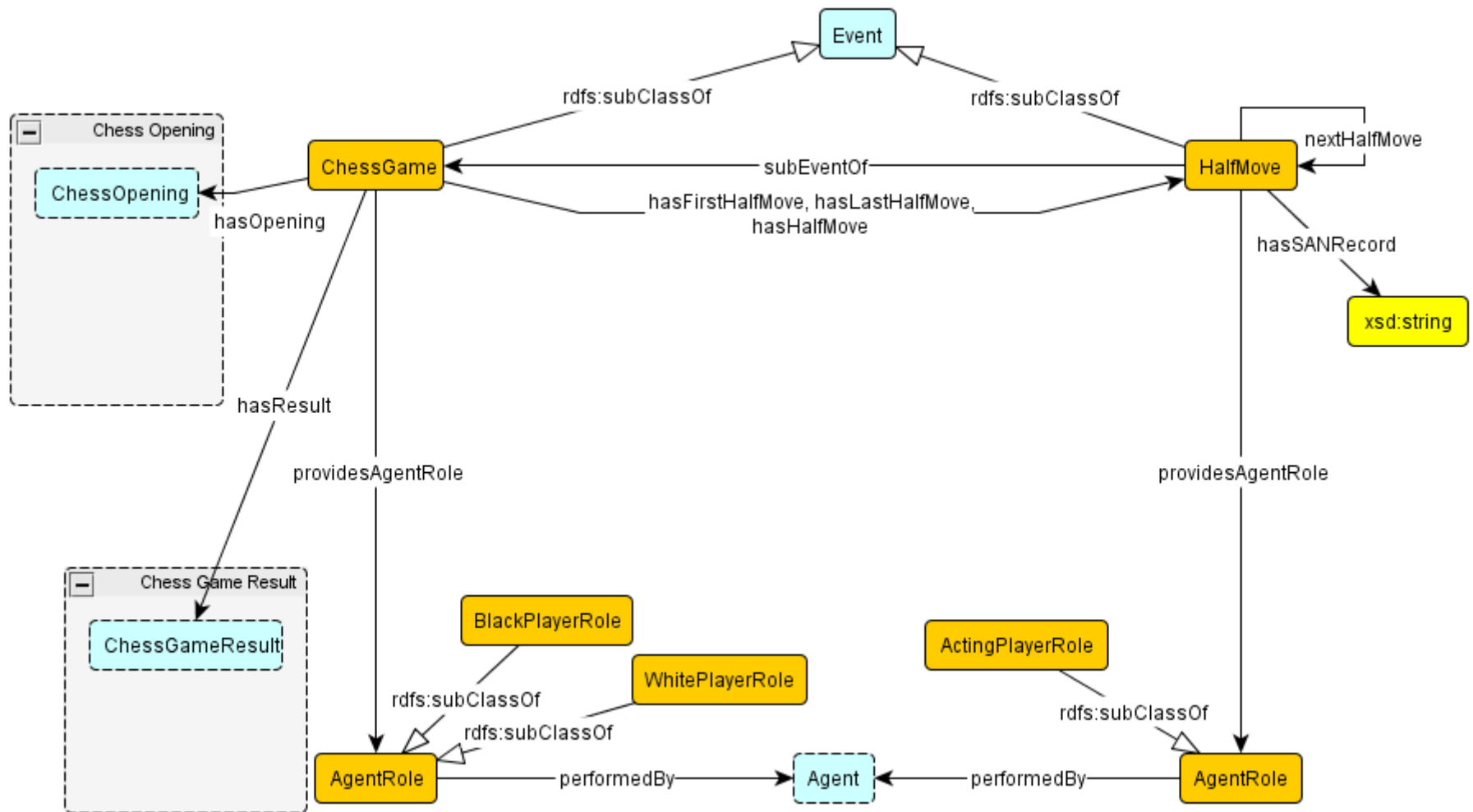


We call these "stubs".

I.e. we're aware that more fine-grained modeling will be needed for some use cases.

But currently there's no reason to do it (not in use case, no data), so we only provide "hooks" for future development of the ontology.

# Commentary and PGN file
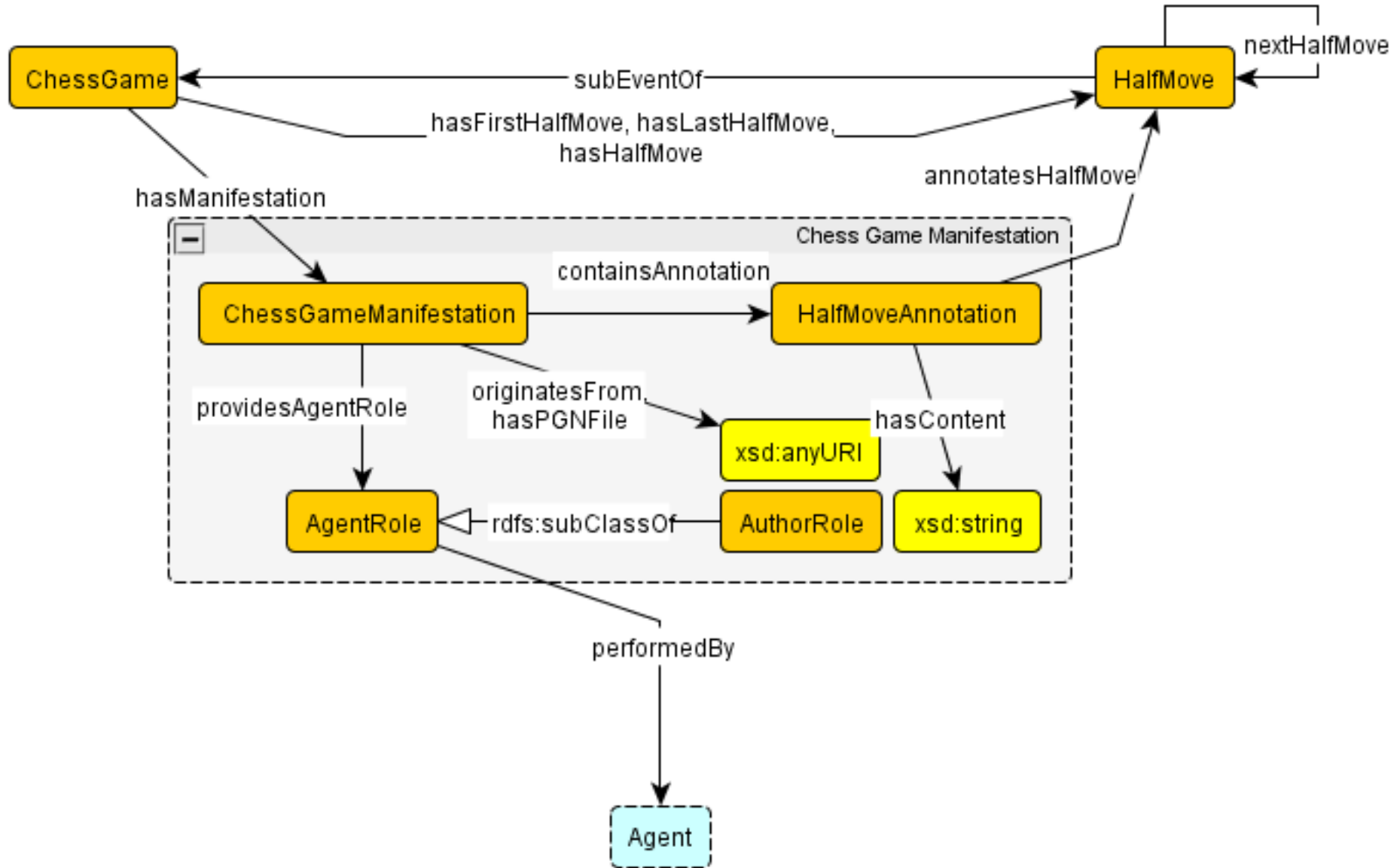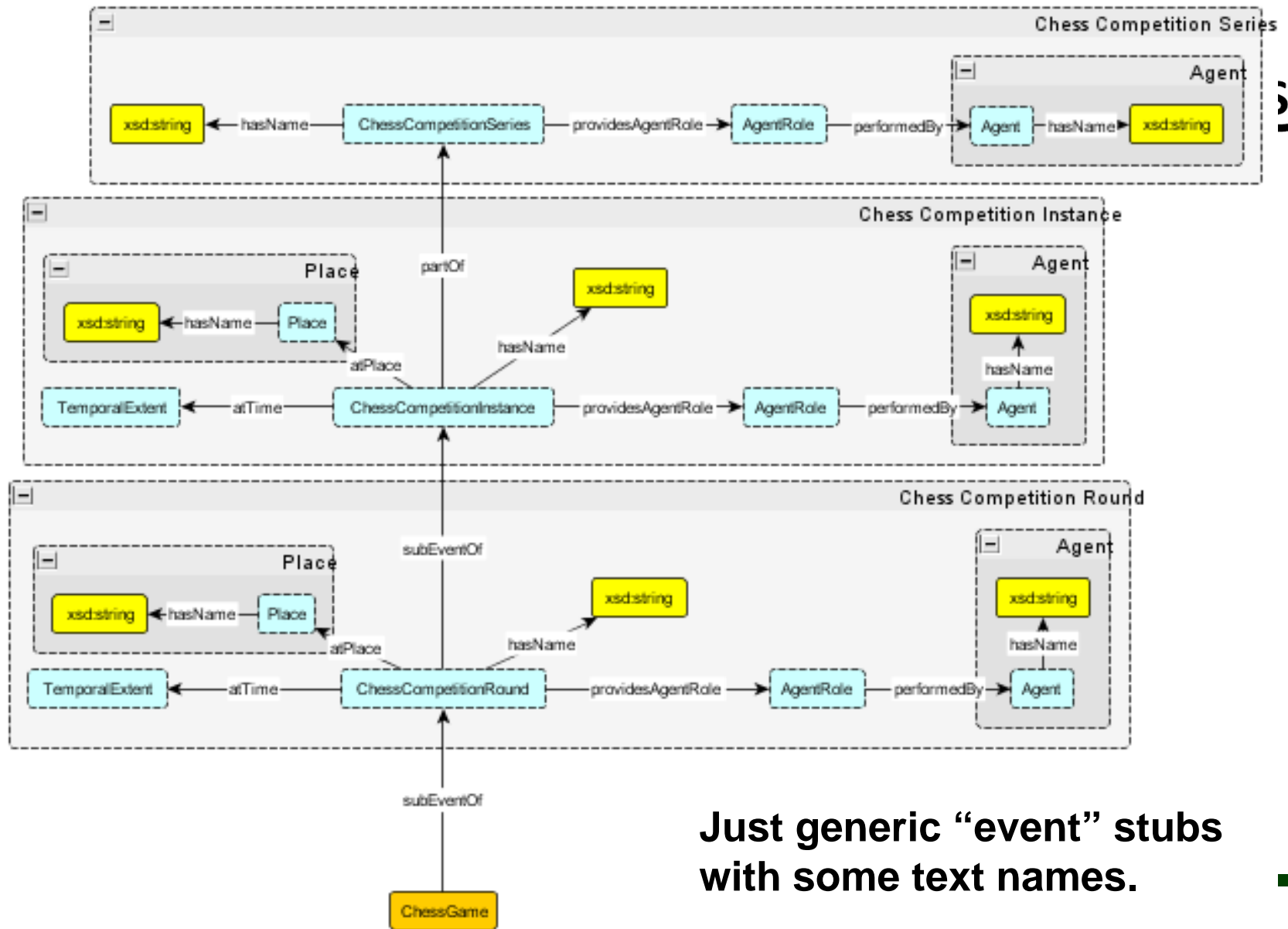
**Just generic "event" stubs with some text names.**

# Adequacy check

- **Triplify sample data** using the ontology.
  Does it work?

- Check if **competency questions** can be answered.

- **Add axioms** as appropriate (the graph is only for intuition, the OWL axioms are the actual ontology).

- (there are more post-hoc details to be taken care of, but let's leave it at that)

# Shortcuts (views)
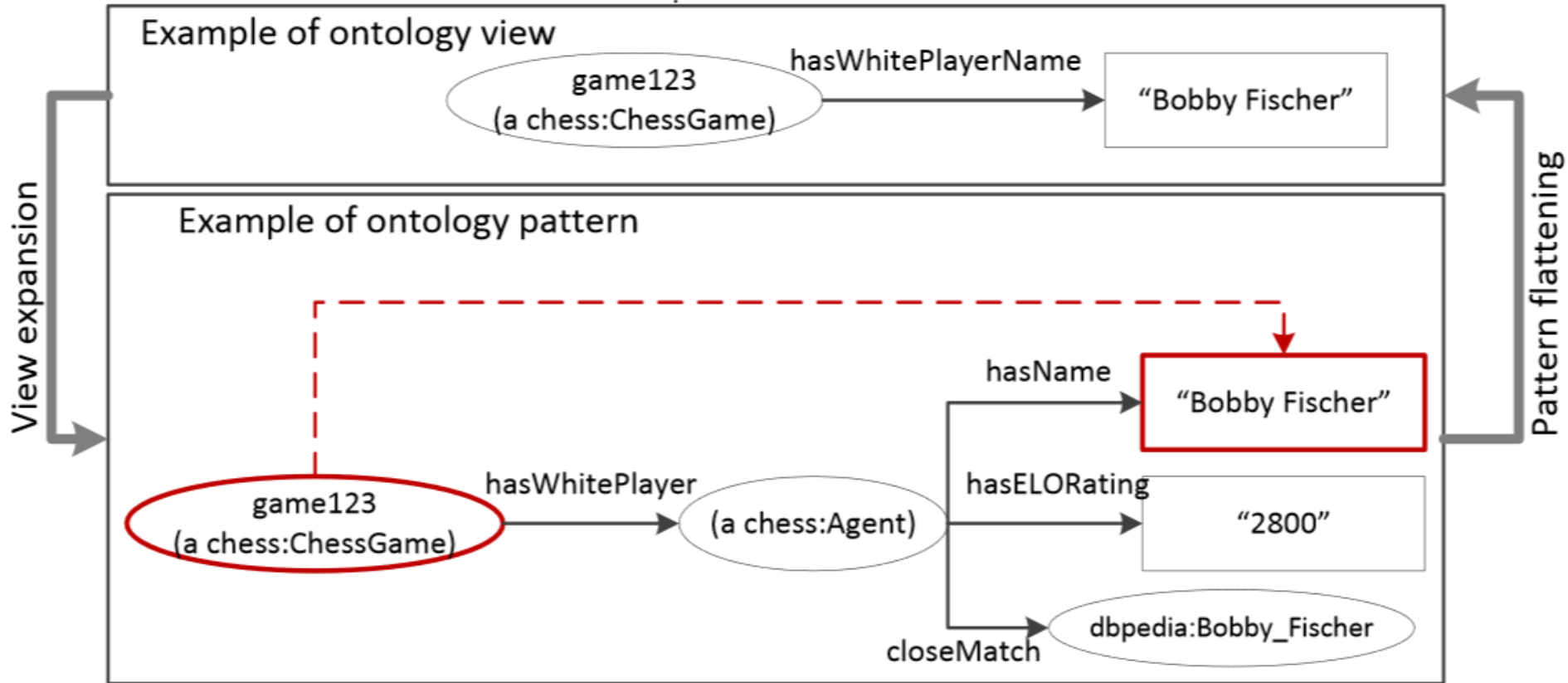
# Modular modeling

Note the modular modeling.

We find that it helps tremendously to

- Focus on a single notion at a time.

- Discuss with domain experts on their grounds without the need to get into technical details.

- Relate to existing ontology design patterns, which helps with reuse and with quality modeling.

# Take Home Message

- **Ontologies cannot be avoided: There is always a conceptual model, even if it's not explicated.**

- **Modular and thorough modeling makes reuse of linked data considerably easier.**

**"… I suppose if we are still seeing a lack of contributions and applications with respect to consuming Linked Data, one would have to question how Linked Data is published in the first-place. And indeed getting publishers to agree on patterns – not just vocabulary – would seem to make consumers' lives that little bit easier. ... Maybe agreement on patterns – not just vocabulary – is what we need to help kick-start consumption of Linked Data"**

# References

- **Pascal Hitzler, Krzysztof Janowicz, Adila Krisnadhi, Ontology modeling with domain experts: The GeoVoCamp experience. In: Proceedings Diversity++ workshop at ISWC 2015. To appear.**

- **Adila A. Krisnadhi, Víctor Rodríguez Doncel, Pascal Hitzler, Michelle Cheatham, Nazifa Karima, Reihaneh Amini, Ashley Coleman, An Ontology Design Pattern for Chess Games. In: Proceedings WOP 2015. To appear.**

- **Víctor Rodríguez-Doncel, Adila A. Krisnadhi, Pascal Hitzler, Michelle Cheatham, Nazifa Karima, Reihaneh Amini, Pattern-Based Linked Data Publication: The Linked Chess Dataset Case. In: Proceedings COLD 2015. To appear.**

- **Adila A. Krisnadhi, Robert Arko, Cynthia Chandler, Michelle Cheatham, Pascal Hitzler, Yingjie Hu, Krzysztof Janowicz, Peng Ji, Nazifa Karima, Adam Shepherd, Peter Wiebe, R2R+BCO-DMO - Linked Oceanographic Datasets. In: Proceedings Diversity++ workshop at ISWC 2015.**

# References

- **Krzysztof Janowicz, Frank van Harmelen, James A. Hendler, Pascal Hitzler, Why the Data Train Needs Semantic Rails. AI Magazine 26 (1), 2015, 5-14.**

- **Krzysztof Janowicz, Pascal Hitzler, Benjamin Adams, Dave Kolas, Charles Vardeman II, Five Stars of Linked Data Vocabulary Use. Semantic Web 5 (3), 2014, 173-176.**

- **Adila A. Krisnadhi, Pascal Hitzler, Krzysztof Janowicz, On capabilities and limitations of OWL regarding typecasting and ontology design pattern views. In: Proceedings OWLED 2015. To appear.**

# Linked Chess Data

See [http://salonica.dia.fi.upm.es:8080/rdfchess/](http://salonica.dia.fi.upm.es:8080/rdfchess/)

(should be available later on [http://chessdata.org](http://chessdata.org))